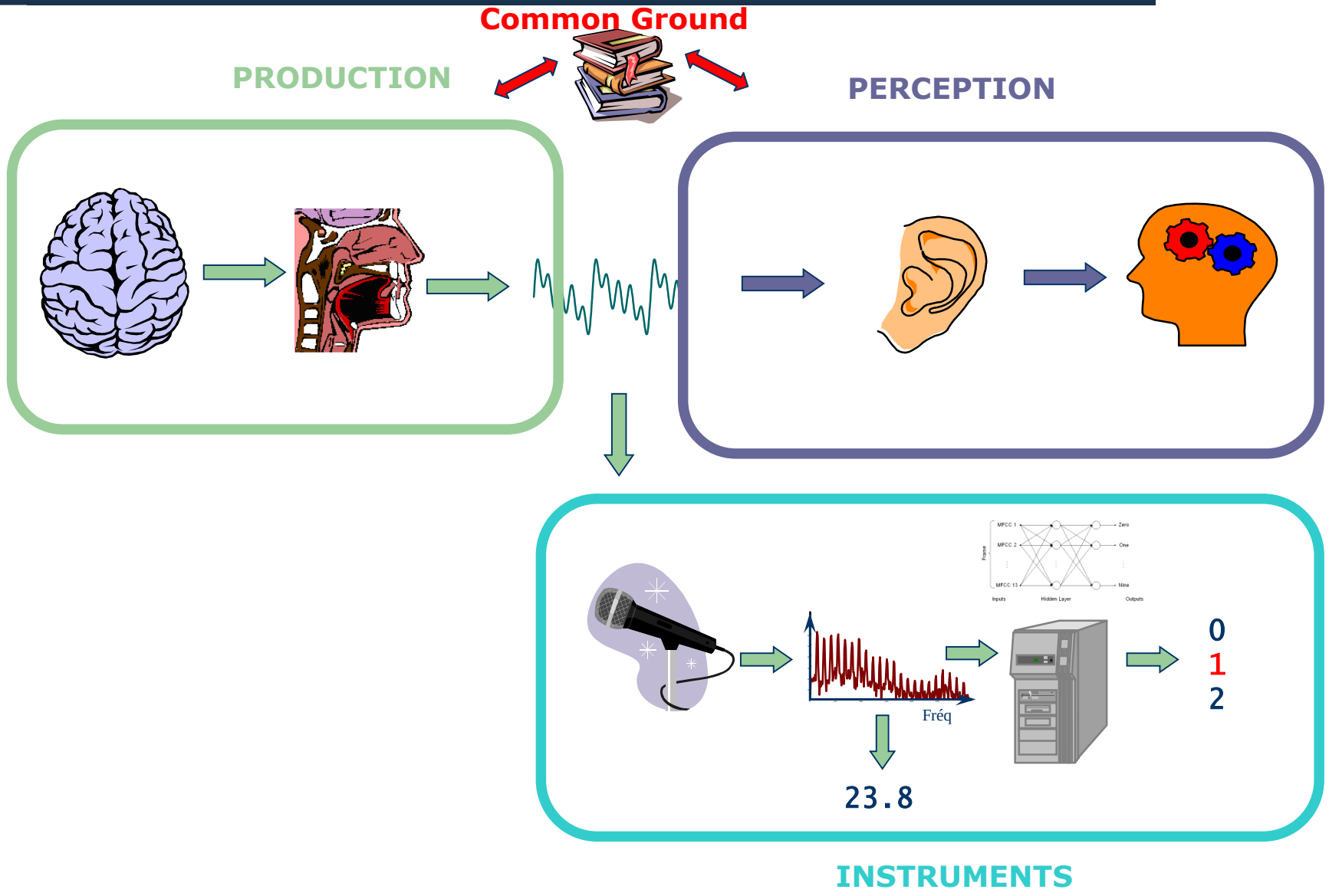
Automatic speech recognition in Laryngology & Phoniatrics practice

Alain Ghio
Laboratoire Parole et Langage
Aix-Marseille University & CNRS
France

CEORL 2019
Bruxelles

Spoken communication



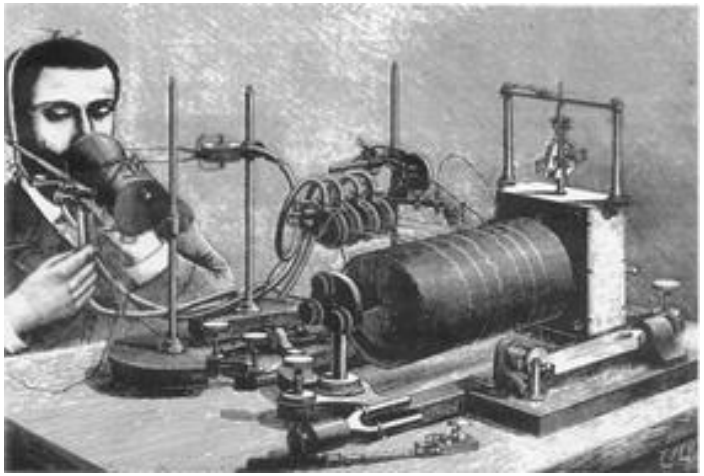
An old story

- 1875 : Société de Linguistique de Paris consults Etienne-Jules Marey (French physiologist who studied movements of animals and humans) in order to…

  « apply the graphic method to the study of the complex and varied movements that occur in speech [...] to obtain an objective trace of the movements of articulatory organs, rib cage, larynx, tongue, lips, soft palate, during the articulation of different phonetic unit»

Because « The ear, the sight and the palpation are not sufficient to apprehend the successive and simultaneous movements necessary for the realization of a phoneme »

- Devices developed by Abbé Rousselot



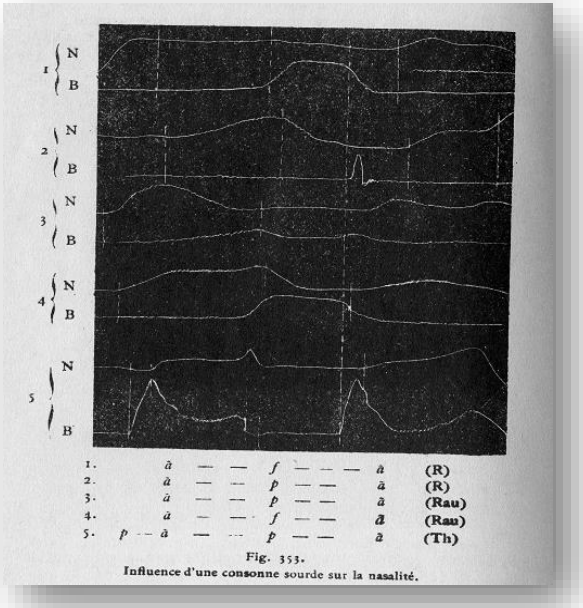


Fig. 353.
Influence d'une consonne sourde sur la nasalité.

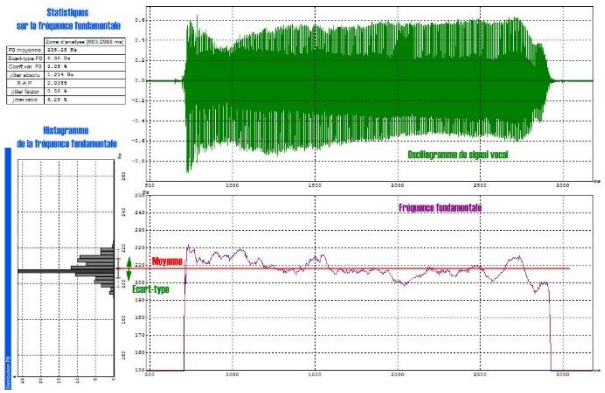
Instrumental assessment

- A basic protocol for functional assessment of voice pathology… (2001), Dejonckere et al., Committee on Phoniatrics of the European Laryngological Society (ELS). Eur Arch Otorhinolaryngol. 2001 Feb;258(2):77-82.

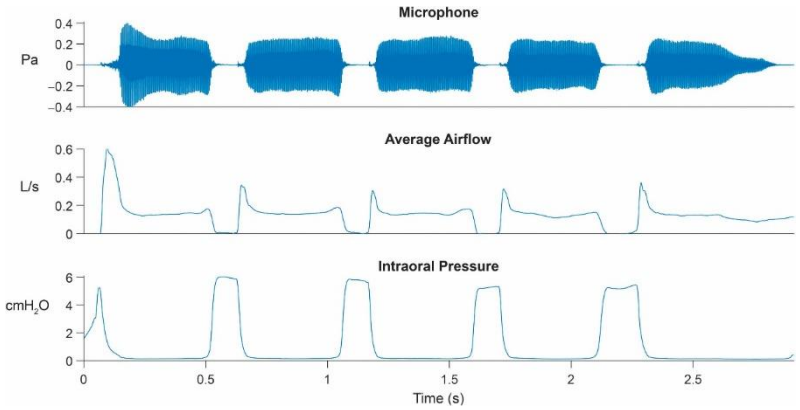
- Recommended Protocols for Instrumental Assessment of Voice: American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function (2018), Patel et al. Am J Speech Lang Pathol. 2018 Aug 6;27(3):887-905

- International consensus (ICON) on basic voice assessment for unilateral vocal fold paralysis, (2018), Mattei et a, European Annals of ORL, Head and Neck Diseases, Volume 135

- Speech disorders assessment less standardized
  - ✓ Motor disorders
  - ✓ "organic" disorders

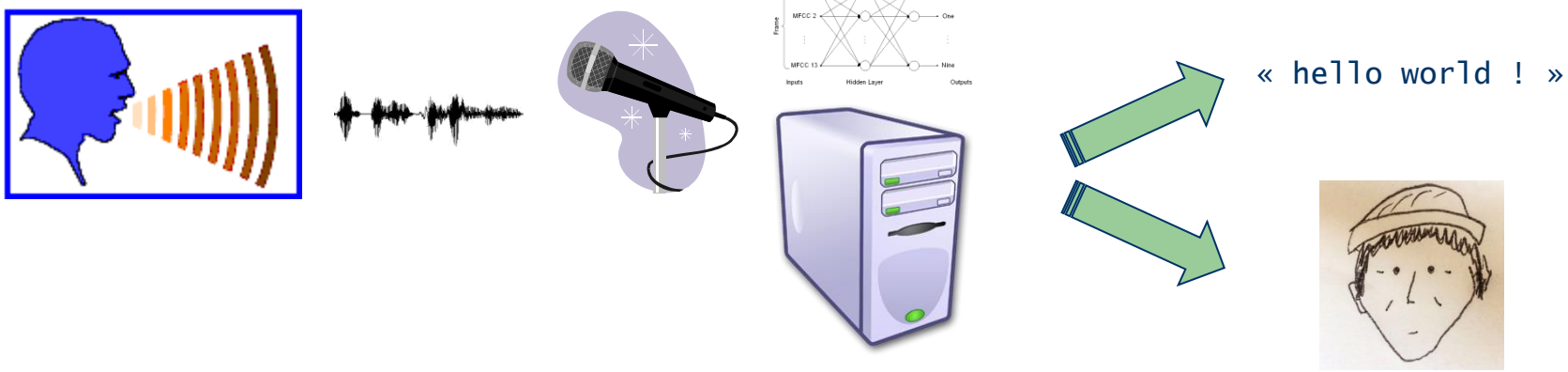


Ghio (2007) : L'évaluation acoustique, Les dysarthries, Solal



Patel et al. (2018) :Protocols for Voice Assessment

What about Automatic Recognition ?



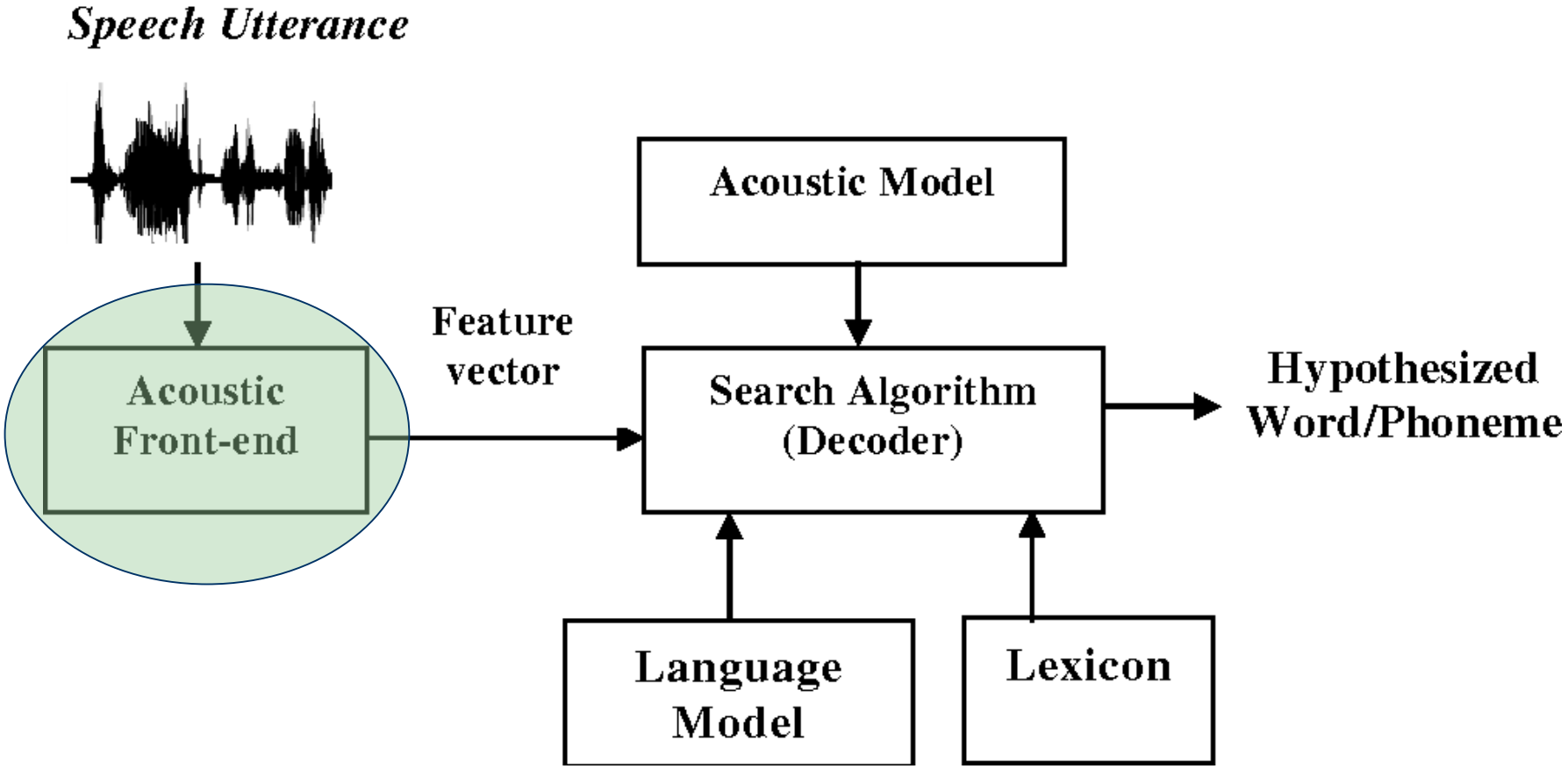
- Automatic Speech Recognition
  - ✓ Decoding of speech content
  - ✓ More adapted to intelligibility

- Automatic Speaker Recognition
  - ✓ Single speaker identification
  - ✓ Voice identification
  - ✓ Can manage groups of speakers (ex: G0, G1, G2, G3)
  - ✓ More adapted to voice disorders

Why automatic recognition is attractive ?

- Simple
  - ✓ A single microphone
  - ✓ Automatic
- Continuous speech
- Can manage complex relation between
  - ✓ Input (speaker, speech)
  - ✓ Output (category)

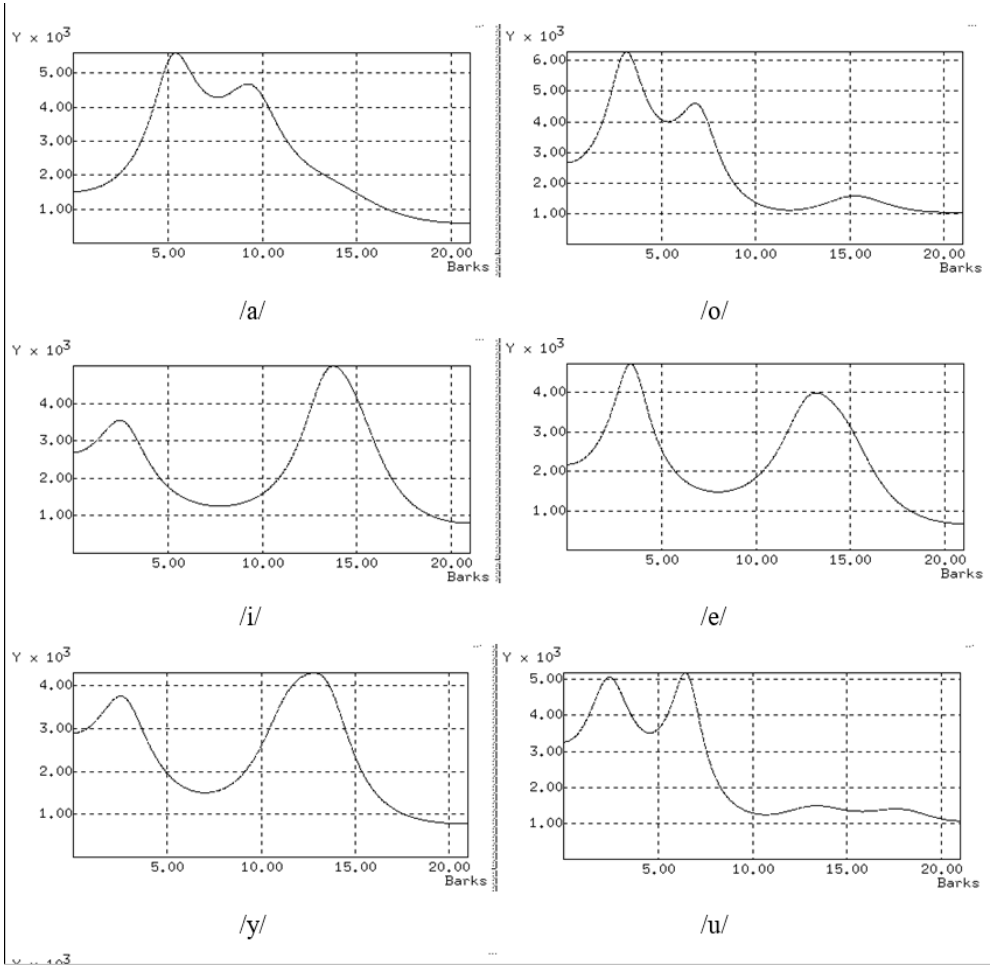
How ?

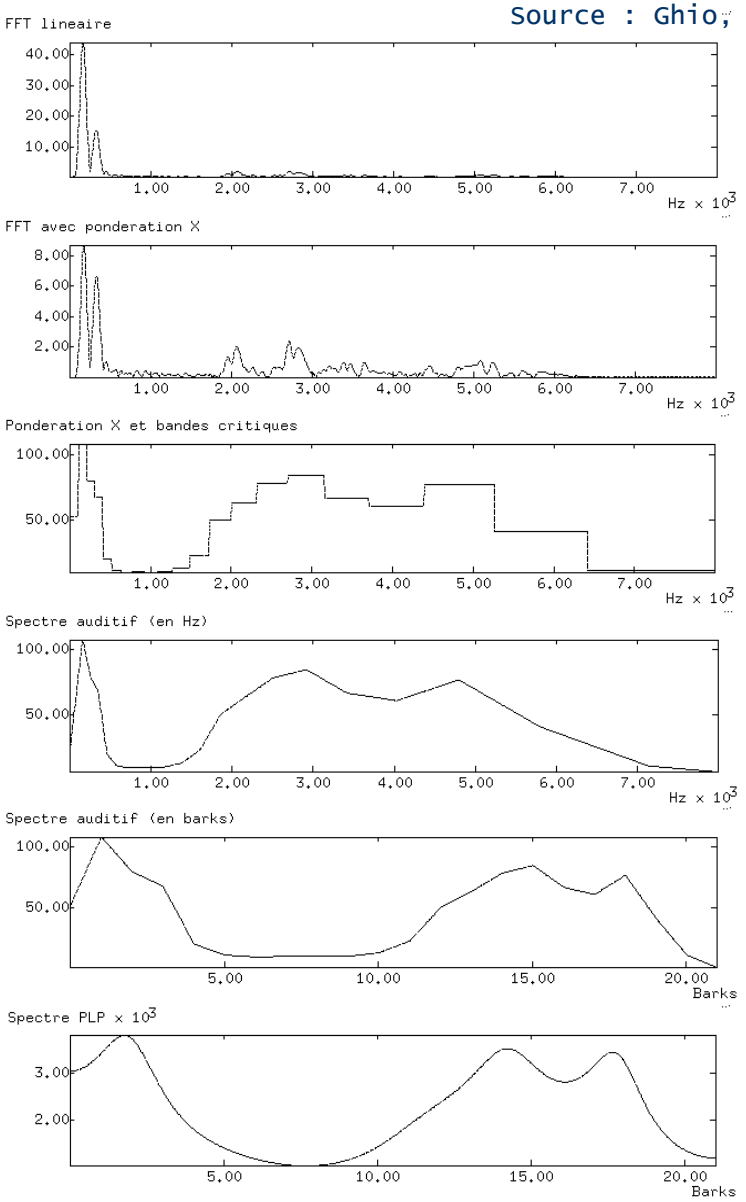


A Review on Automatic Speech Recognition Architecture and Approaches, Karpagavalli et al. (2016)

Acoustic front end ?

Source : Ghio, 1997

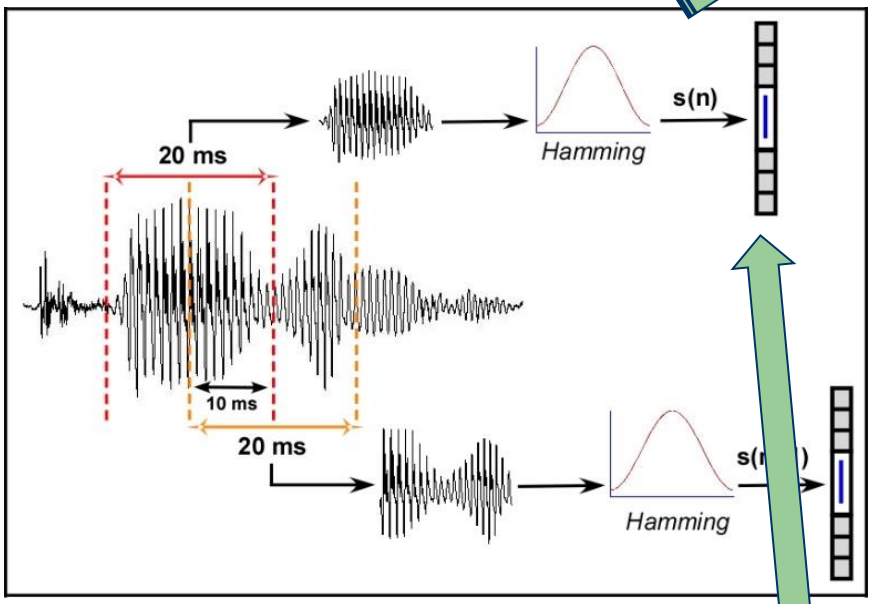


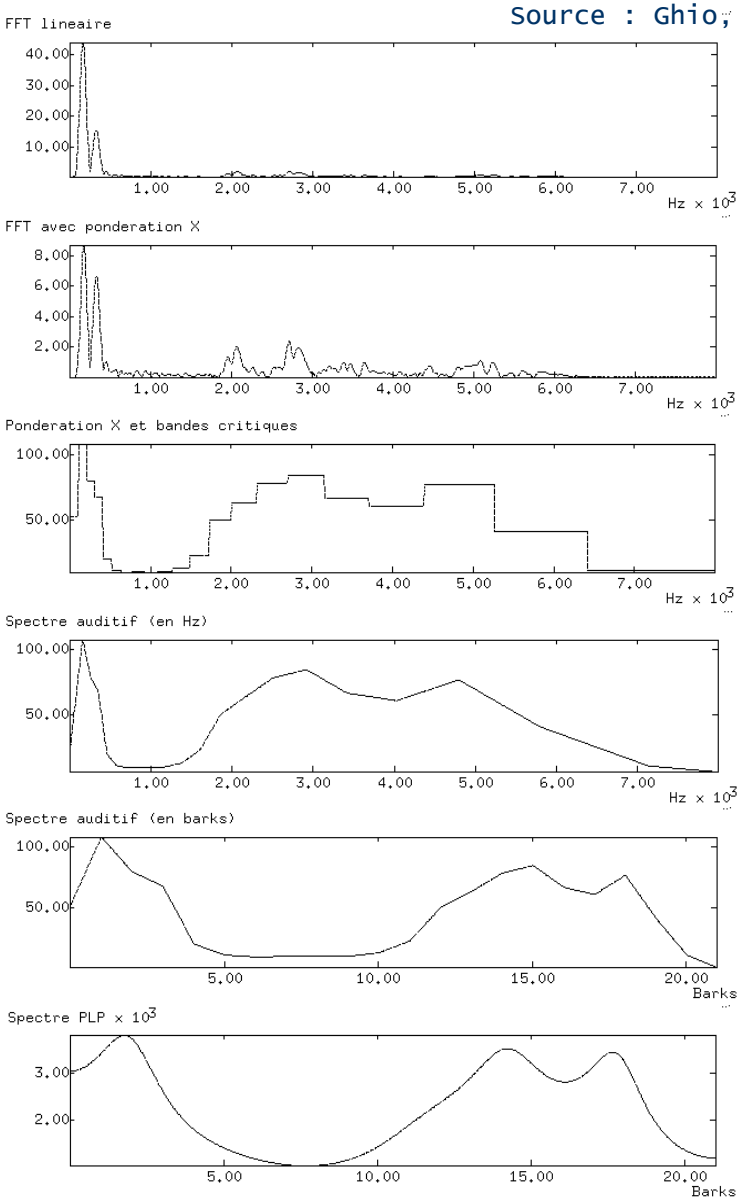


This curve can be modelized
by 10 coefficients
= 1 vector

Acoustic front end ?

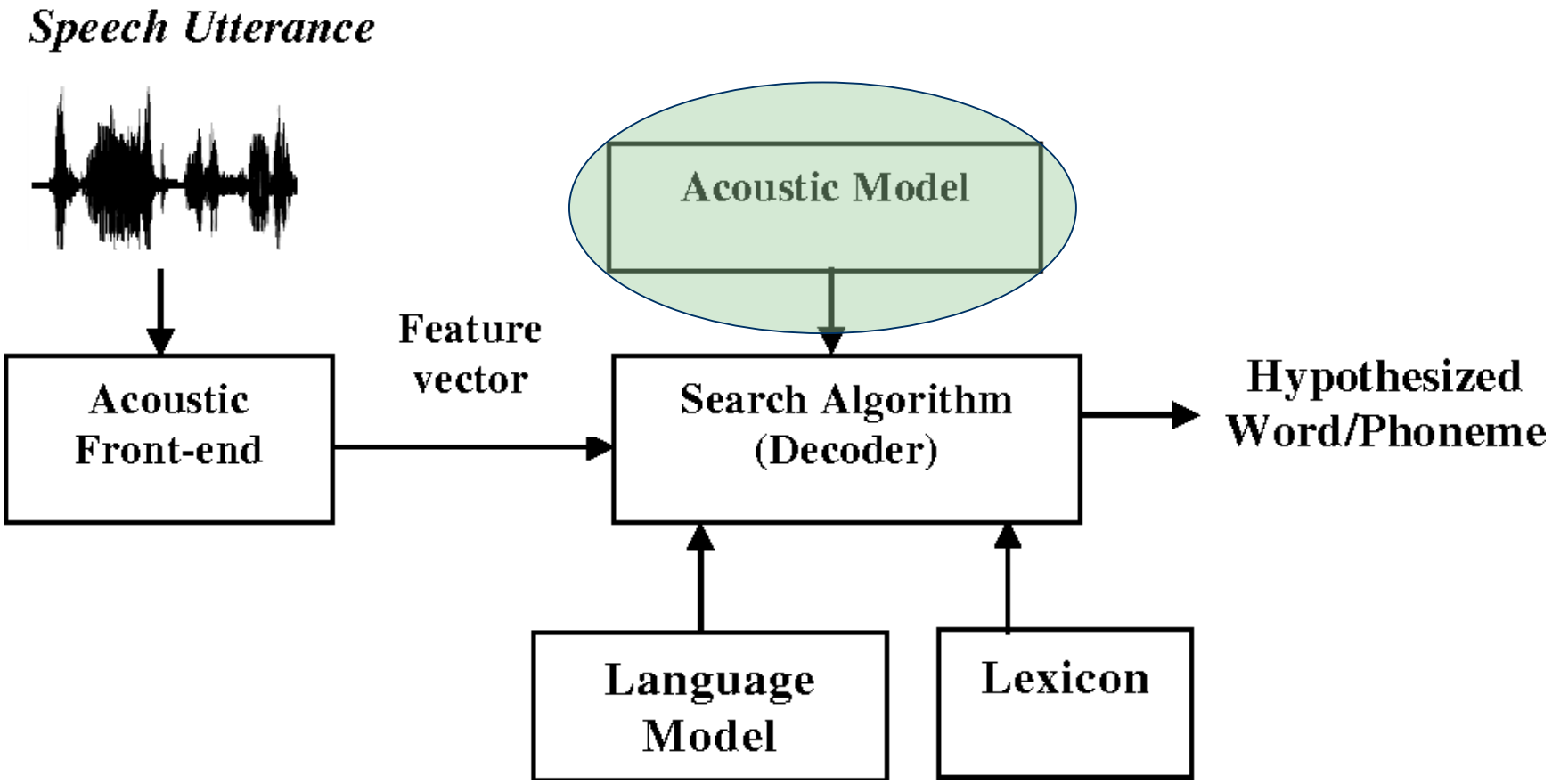
Source : Ghio, 1997





This curve can be modelized
by 10 coefficients
= 1 vector

How ?



A Review on Automatic Speech Recognition Architecture and Approaches, Karpagavalli et al. (2016)

Modelling Acoustic ?

- Modelling acoustic
  - ✓ Based on training data (available corpus)
- Phoneme modelling
  - ✓ Mixture of all different phonemes available on the corpus
  - ✓ Phonemes models are based on the corpus
  - ✓ Sensitivity to the training corpus
- Speaker(s) modelling
  - ✓ Mixture of all sounds of a speaker or an homogeneous group of speakers (ex: slight dysphonia)

News : deep learning

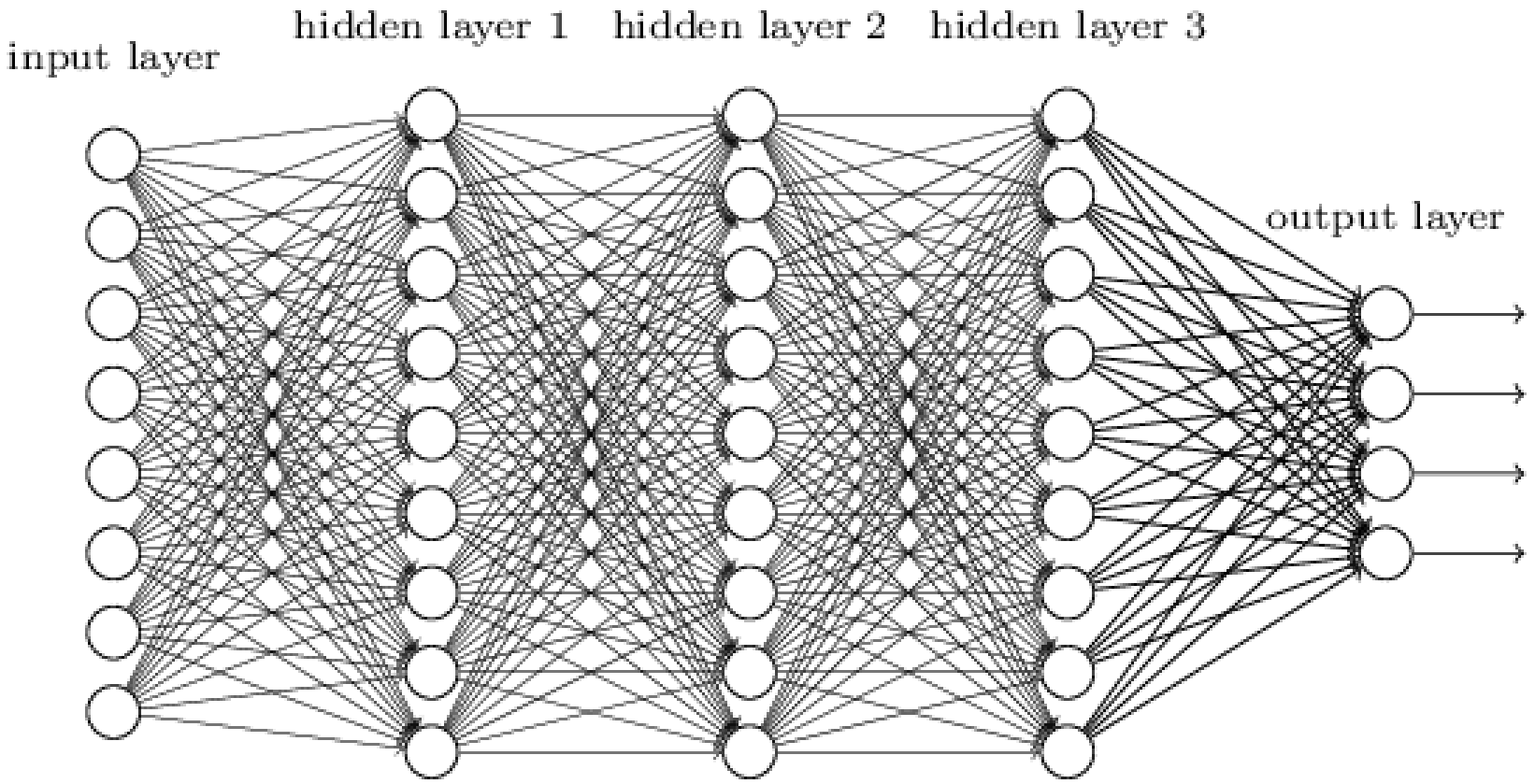
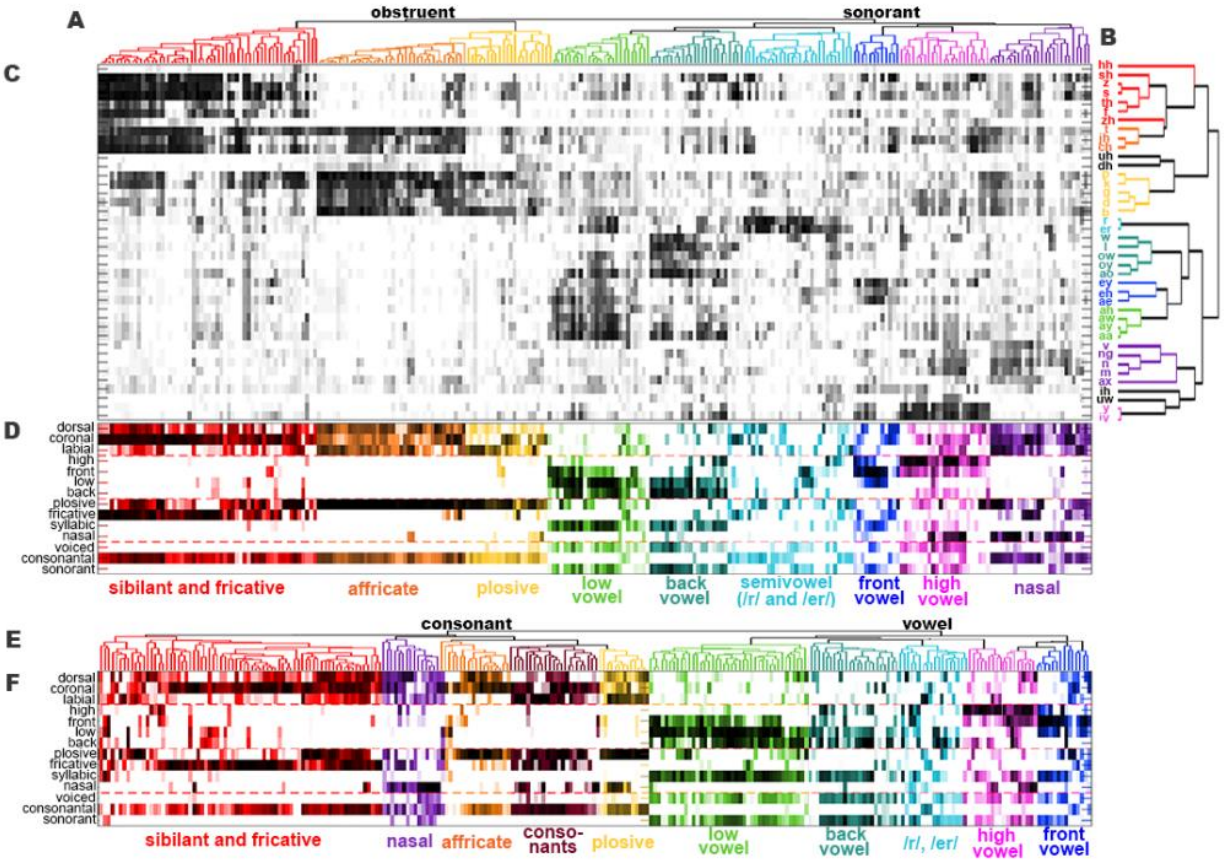


Image Source: https://www.quora.com/In-multilayer-neural-networks-are-weights-in-hidden-layers-closer-to-the-input-updated-more-strongly-than-weights-in-layers-closer-to-the-output

NN and speech

- Nagamine et al. Exploring How Deep Neural Networks Form Phonemic Categories, INTERSPEECH 2015



ASR in Laryngology & Phoniatrics

- Speaker categorization
  - ✓ Train ASR system with groups of speakers
  - ✓ Compare a new patient with the ASR
    - ✧ *M. Wester, Automatic classification of voice quality:Comparing regression models and hidden markov models,in: VOICEDATA98, Symposium on Databases in VoiceQuality Research and Education, 1998, pp. 92–97*
    - ✧ *Fredouille, Pouchoulin, Bonastre, Azzarello, Giovanni, Ghio. Application of Automatic Speaker Recognition techniques to pathological voice assessment (dysphonia). Interspeech, 2005, Lisboa, France. pp.149-152.*

	Classification system Response			
Test Gr.	0	1	2	3
0	**19**	1	0	0
1	3	**14**	2	1
2	2	7	**9**	2
3	0	0	7	**13**

- Need a labeled training corpus
  - ✓ Dependent of primary classification of speakers based on… perception which is not perfectly reliable
- Circularity
- Need a real gold standard

ASR in Laryngology & Phoniatrics

- Speech distorsion categorization
  - ✓ Train ASR system with « normal » speech
  - ✓ Compare speech « transcription » of new patient by ASR
    - *Maier, Schuster, Batliner, Nöth, et al. "Automatic scoring of the intelligibility in patients with cancer of the oral cavity", Interspeech 2007, 1206-1209.*
    - *Middag, C., Van Nuffelen, G., Martens, J.P., De Bodt, M., 2008. Objective intelligibility assessment of pathological speakers. In: Proceedings ofthe International Conference on Spoken Language Processing, Brisbane, Australia, pp. 1745–1748.*
- Needs
  - ✓ a huge training corpus
  - ✓ a lexicon and language model
- Acoustic-phonetic decoding or speech decoding

Conclusion

- We need to validate Automatic speech recognition techniques in Laryngology & Phoniatrics context
- Comparison with human perception but a gold standard is needed
- Accepted if not a blackbox
- Useful if we can improve our knowledge
  - ✓ Deep Neural Network could be a good candidate

Thank you

Research on ASR with

Gilles Pouchoulin (LPL, Univ. Aix Marseille)

Corinne Fredouille (Computer Sciences, Univ. Avignon)

Jérome Farinas (Computer Sciences, Univ. Toulouse)